Cas d’ús i guia de resolució en R

Cadenes de Markov: estudi del vent

PID_00262438

Francesc Pozo Montero
Jordi Ripoll Missé

Universitat
Oberta
de Catalunya

Francesc Pozo Montero

Llicenciat en Matemàtiques per la Universitat de Barcelona (2000) i doctor en Matemàtica Aplicada per la Universitat Politècnica de Catalunya (2005). Ha estat professor associat a la Universitat Autònoma de Barcelona i professor associat, col·laborador i actualment professor agregat a la Universitat Politècnica de Catalunya. A més, és cofundador del Grup d'Innovació Matemàtica E-learning (GIMEL), responsable de diversos projectes d'innovació docent i autor de diverses publicacions. Com a membre del grup de recerca consolidat CoDAlab, centra la recerca en la teoria de control i les aplicacions en enginyeria mecànica i civil, com també en l'ús de la ciència de dades per al monitoratge de la integritat estructural i per al monitoratge de la condició, sobretot en turbines eòliques.

Jordi Ripoll Missé

Llicenciat en Matemàtiques i doctor en Ciències Matemàtiques per la Universitat de Barcelona (2005). Professor col·laborador de la Universitat Oberta de Catalunya des del 2011 i professor del Departament d'Informàtica, Matemàtica Aplicada i Estadística de la Universitat de Girona (UdG) des del 1996, on actualment és professor agregat i desenvolupa tasques de recerca en l'àmbit de la biologia matemàtica (models amb equacions en derivades parcials i dinàmica evolutiva). També ha estat professor i tutor de la UNED en dues etapes, primer al centre associat de Terrassa i actualment al de Girona. Ha participat en nombrosos projectes d'innovació docent, especialment pel que fa a l'aprenentatge de les matemàtiques en línia.

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats per la professora: Cristina Cano Bastidas (2019)

Primera edició: febrer 2019

Av. Tibidabo, 39-43, 08035 Barcelona
Disseny: Manel Andreu
Realització editorial: Oberta UOC Publishing, SL

Índex

1. Cas d'ús: estudi dels canvis de direcció del vent

En aquest cas d'ús es vol exemplificar, a partir d'unes dades reals (mesures empíriques) d'una cadena de Markov per a les direccions del vent en una localització, com es pot deduir la distribució de probabilitat de la direcció del vent en aquella localització geogràfica.

La direcció a la qual bufa el vent en un lloc determinat es pot considerar com un procés estocàstic en el qual en tenim prou de saber quines són les probabilitats que el vent canviï de direcció en cada període de temps.

Segons les dades recopilades en la tesi doctoral de Palafox Duarte i de la inferència estadística feta en aquell treball, podem considerar una cadena de Markov homogènia per a la direcció del vent a la ciutat de Guanajuato (Mèxic, latitud 21°00′59″N), per a un temps discret mesurat en hores. La direcció del vent, és a dir, l'orientació del vector del vent, es defineix com la direcció i el sentit des del qual bufa el vent. Les direccions del vent s'han classificat dividint la circumferència en vuit sectors iguals de 45° i associant a cada sector un dels punts cardinals principals i secundaris:

Bibliografia recomanada

Martha Cecilia Palafox Duarte (2009). *Inferencia estadística para cadenas de Markov*.

$$
\begin{array}{ccccc}
\mathrm{NW} & & \mathrm{N} & & \mathrm{NE} \\
 & \nwarrow & \uparrow & \nearrow & \\
\mathrm{W} & \longleftarrow & \bullet & \longrightarrow & \mathrm{E} \\
 & \swarrow & \downarrow & \searrow & \\
\mathrm{SW} & & \mathrm{S} & & \mathrm{SE}
\end{array}
$$

Per tant, aquest model té vuit estats possibles etiquetats pels punts cardinals

$$\{\mathrm{N}, \mathrm{NE}, \mathrm{E}, \mathrm{SE}, \mathrm{S}, \mathrm{SW}, \mathrm{W}, \mathrm{NW}\}$$

Les probabilitats de canviar cada hora d'una direcció a una altra són les que dona la matriu següent:

$$P = \begin{array}{cc} \begin{array}{cccccccc} \text{N} & \text{NE} & \text{E} & \text{SE} & \text{S} & \text{SW} & \text{W} & \text{NW} \end{array} & \\ \begin{pmatrix} 0.541 & 0.3604 & 0.261 & 0.18 & 0.12 & 0.16 & 0.37 & 0.36 \\ 0.311 & 0.44 & 0.281 & 0.11 & 0.070 & 0.057 & 0.13 & 0.206 \\ 0.061 & 0.11 & 0.27 & 0.102 & 0.031 & 0.023 & 0.06 & 0.033 \\ 0.034 & 0.041 & 0.083 & 0.311 & 0.22 & 0.034 & 0.02 & 0.236 \\ 0.021 & 0.0264 & 0.06 & 0.181 & 0.323 & 0.256 & 0.1 & 0.066 \\ 0.011 & 0.007 & 0.03 & 0.081 & 0.16 & 0.27 & 0.12 & 0 \\ 0.008 & 0.0062 & 0.01 & 0.035 & 0.07 & 0.12 & 0.14 & 0.033 \\ 0.013 & 0.009 & 0.005 & 0 & 0.006 & 0.08 & 0.06 & 0.066 \end{pmatrix} & \begin{array}{c} \text{N} \\ \text{NE} \\ \text{E} \\ \text{SE} \\ \text{S} \\ \text{SW} \\ \text{W} \\ \text{NW} \end{array} \end{array}$$

en què cada element P_{ij} representa la probabilitat de canviar de la direcció j a la direcció i.

Val la pena destacar que l'anàlisi que hem fet aquí sobre els canvis en la direcció del vent en una ciutat específica, és a dir, la recopilació de dades (cadena de canvis del vent per hores), la inferència estadística per estimar la matriu de probabilitats de transició i la predicció posterior de la distribució del vent, pot ser útil per fer una planificació urbanística (instal·lacions esportives a l'aire lliure, zones verdes...), per construir parcs eòlics, etc.

Quan calgui, per mitjà del llenguatge de programació R resoleu aquestes qüestions:

1. $P \geq 0$ és una matriu simètrica? Quins elements de la matriu P són zero?

2. En quina direcció hi ha un probabilitat més alta que el vent continuï bufant en la mateixa direcció? I en quina hi ha la probabilitat més alta que canviï de direcció?

3. Calculeu la matriu P^2 i comproveu que $P^2 > 0$. Segons aquest model, si en un determinat instant de temps observem que el vent té direcció sud-est (SE), quina és la probabilitat que el vent canviï en direcció nord-oest (NW) al cap de dos períodes de temps? I de NW a SW?

4. Trobeu els valors propis λ_i, $i = 1 \ldots 8$, de la matriu P. Ordeneu-los de gran a petit segons el seu mòdul $|\lambda_i|$. Comproveu que es compleix

$$\lambda_1 = 1 > |\lambda_2| \geq \cdots \geq |\lambda_8| \,.$$

5. Calculeu el vector propi $\vec{v}$ de valor propi $\lambda_1 = 1$ de la matriu P. Quina és la distribució de probabilitat estacionària d'aquest model? Normalitzeu el vector propi $\frac{1}{v_1+\cdots+v_n}(v_1,\ldots,v_n)$.

6. Calculeu la matriu P^{40}. Què observeu? Quina relació hi ha amb el vector propi trobat en l'apartat anterior?

7. Finalment, calculeu la mitjana de cada una de les files de la matriu P i escriviu-ne el resultat en forma de vector. Aquest vector obtingut és molt diferent del vector propi (normalitzat) trobat abans? En aquest cas, podeu fer servir `rowMeans(P)`.

2. Solucions de les qüestions

1. Definim la matriu P, de dimensió 8×8, de les probabilitats de canvi de la direcció del vent i observem que tots els elements són estrictament positius, excepte $P_{84} = 0$ i $P_{68} = 0$. Això significa que quan el vent bufa de sud-est no canvia mai directament a nord-oest durant la propera hora i que quan bufa de nord-oest tampoc canvia mai directament a sud-oest, respectivament.

```
> Etiquetes <- c("N", "NE", "E", "SE","S", "SW", "W", "NW")
> P <- matrix(c(0.541 , 0.3604 , 0.261 , 0.18 , 0.12 , 0.16 , 0.37 , 0.36,
+               0.311 , 0.44 , 0.281 , 0.11 , 0.070 , 0.057 , 0.13 , 0.206,
+               0.061 , 0.11 , 0.27 , 0.102 , 0.031 , 0.023 , 0.06 , 0.033,
+               0.034 , 0.041 , 0.083 , 0.311 , 0.22 , 0.034 , 0.02 , 0.236,
+               0.021 , 0.0264 , 0.06 , 0.181 , 0.323 , 0.256 , 0.1 , 0.066,
+               0.011 , 0.007 , 0.03 , 0.081 , 0.16 , 0.27 , 0.12 , 0,
+               0.008 , 0.0062 , 0.01 , 0.035 , 0.07 , 0.12 , 0.14 , 0.033,
+               0.013 , 0.009 , 0.005 , 0 , 0.006 , 0.08 , 0.06 , 0.066), 8,8,
+               byrow=T, dimnames = list(Etiquetes, Etiquetes))
> # install.packages("markovchain")
> library(markovchain)
> # Creem un nou objecte, que és una cadena de Markov (per columnes)
> CM <- new("markovchain", states = Etiquetes, byrow = F,
+           transitionMatrix = P, name = "Canvis de direcció del vent")
> CM

Canvis de direcció del vent
 A  8 - dimensional discrete Markov Chain defined by the following states:
 N, NE, E, SE, S, SW, W, NW
 The transition matrix (by cols) is defined as follows:
       N     NE     E    SE     S    SW    W    NW
N  0.541 0.3604 0.261 0.180 0.120 0.160 0.37 0.360
NE 0.311 0.4400 0.281 0.110 0.070 0.057 0.13 0.206
E  0.061 0.1100 0.270 0.102 0.031 0.023 0.06 0.033
SE 0.034 0.0410 0.083 0.311 0.220 0.034 0.02 0.236
S  0.021 0.0264 0.060 0.181 0.323 0.256 0.10 0.066
SW 0.011 0.0070 0.030 0.081 0.160 0.270 0.12 0.000
W  0.008 0.0062 0.010 0.035 0.070 0.120 0.14 0.033
NW 0.013 0.0090 0.005 0.000 0.006 0.080 0.06 0.066
```

2. Les probabilitats que el vent continuï bufant en la mateixa direcció són els elements de la diagonal de P. Per tant, la probabilitat més alta que el vent vagi bufant en la mateixa direcció (nord, N) és $P_{11} = 54.1\%$. D'altra banda, les probabilitats que el vent canviï de direcció són els elements externs de la diagonal de P. Per tant, la probabilitat més alta que el vent canviï de direcció (de W a N) és $P_{17} = 37\%$.

3. Calculem P^2 i observem que, efectivament, és una matriu amb tots els elements estrictament positius. Això implica que, a partir d'una direcció, podem canviar a qualsevol altra al llarg d'un interval de temps de dues hores (cadena o matriu irreductible). Si el vent té direcció sud-est, la probabilitat que el vent canviï a direcció nord-oest al cap de dues hores és $P^2_{84} = 1.35\%$. En canvi, la probabilitat de canviar de nord-oest a sud-oest al cap de dues hores és $P^2_{68} = 4\%$.

```
> CM^2  # cadena de Markov al cap de dos períodes de temps (2 h)

Canvis de direcció del vent^2
 A  8 - dimensional discrete Markov Chain defined by the following states:
 N, NE, E, SE, S, SW, W, NW
 The transition matrix (by cols) is defined as follows:
           N        NE         E       SE        S         SW        W        NW
N  0.4387264 0.3994644 0.3453834 0.267256 0.230259 0.2663458 0.370882 0.3639854
NE 0.3317870 0.3460114 0.2980510 0.189089 0.143097 0.1504330 0.235730 0.2603390
E  0.0889620 0.1059148 0.1315120 0.091916 0.063921 0.0496940 0.071350 0.0838060
SE 0.0550040 0.0604686 0.0842180 0.159091 0.157259 0.1070400 0.072150 0.1275770
S  0.0406424 0.0447386 0.0725124 0.151794 0.198733 0.1814868 0.099402 0.0866684
SW 0.0200020 0.0205234 0.0385610 0.086031 0.123840 0.1338630 0.073600 0.0400280
W  0.0123952 0.0119992 0.0188002 0.041317 0.061212 0.0728134 0.048046 0.0241652
NW 0.0124810 0.0108796 0.0109620 0.013506 0.021679 0.0383240 0.028840 0.0134310
```

4. Com que $P \geq 0$ és una matriu positiva, sempre té com a mínim un valor propi positiu $\lambda \geq 0$. A més, com que es tracta d'una matriu estocàstica (cadena de Markov), el valor propi dominant de P sempre és $\lambda_1 = 1$. Podem calcular, també, tots els valors propis de la matriu P, ordenats de gran a petit segons el seu mòdul:

$$\lambda_1 = 1 > |\lambda_2| = 0.58 > |\lambda_3| = 0.26 > |\lambda_4| = 0.17 > |\lambda_5| = 0.13 >$$
$$> |\lambda_6| = 0.11 > |\lambda_7| = 0.05 = |\lambda_8| = 0.05$$

Si es compleix aquesta última propietat, direm que $\lambda_1 = 1$ és el valor propi estrictament dominant de la matriu.

```
> V <- eigen(P)
> vaps <- V$values # valors propis de la matriu
> abs(vaps) # valors propis en mòdul

[1] 1.00000000 0.58015192 0.26267674 0.16790667 0.13261832 0.11360841 0.05325669
[8] 0.05325669

> abs(vaps)[1] > abs(vaps)[2] # comprovació

[1] TRUE
```

5. En aquest exemple podem veure que hi ha la distribució límit d'estats (direccions del vent) i coincideix amb la distribució estacionària. L'estat estacionari és el vector propi $\vec{v} \geq 0$ de valor propi $\lambda_1 = 1$ de la matriu P. Si calculem aquest vector i el normalitzem de manera que la suma de les components sigui 1, obtindrem aquesta distribució de probabilitat per a la direcció del vent:

N	NE	E	SE	S	SW	W	NW
38%	29.7%	9.4%	7.9%	7.3%	4.1%	2.2%	1.4%

Aquesta distribució indica que, en aquesta localització, el vent bufarà majoritàriament de component nord o nord-est.

```
> # Estat estacionari = distribució límit d'estats
> veps <- V$vectors # vectors propis de la matriu
> steadystate <- abs(veps[,1])
> (steadystate <- steadystate/sum(steadystate))
```

```
[1] 0.38002783 0.29737390 0.09436523 0.07852117 0.07278991 0.04075919 0.02212273
[8] 0.01404003
```

6. Calculem la potència 40 de la matriu P i observem que totes les columnes d'aquesta nova matriu coincideixen (aproximadament) amb el vector propi normalitzat $\vec{v}$. De fet, les successives potències, $P^2, P^3, \ldots$, tendeixen cap al mateix resultat.

```
> CM^40  # cadena de Markov al cap de quaranta períodes de temps (40 h)
```

```
Canvis de direcció del vent^40
 A  8 - dimensional discrete Markov Chain defined by the following states:
 N, NE, E, SE, S, SW, W, NW
 The transition matrix (by cols) is defined as follows:
            N         NE          E         SE          S         SW          W
N  0.38002783 0.38002783 0.38002783 0.38002783 0.38002783 0.38002783 0.38002783
NE 0.29737390 0.29737390 0.29737390 0.29737390 0.29737390 0.29737390 0.29737390
E  0.09436523 0.09436523 0.09436523 0.09436523 0.09436523 0.09436523 0.09436523
SE 0.07852117 0.07852117 0.07852117 0.07852117 0.07852117 0.07852117 0.07852117
S  0.07278991 0.07278991 0.07278991 0.07278991 0.07278991 0.07278991 0.07278991
SW 0.04075919 0.04075919 0.04075919 0.04075919 0.04075919 0.04075919 0.04075919
W  0.02212273 0.02212273 0.02212273 0.02212273 0.02212273 0.02212273 0.02212273
NW 0.01404003 0.01404003 0.01404003 0.01404003 0.01404003 0.01404003 0.01404003
           NW
N  0.38002783
NE 0.29737390
E  0.09436523
SE 0.07852117
S  0.07278991
SW 0.04075919
```

```
W  0.02212273
NW 0.01404003
```

7. Si calculem la mitjana per files de la matriu P, obtindrem una primera aproximació a la distribució estacionària del model:

	N	NE	E	SE	S	SW	W	NW
VEP	38%	29.7%	9.4%	7.9%	7.3%	4.1%	2.2%	1.4%
mitjana files	29.4%	20.1%	8.6%	12.2%	12.9%	8.5%	5.3%	3.0%

Podem observar que aquests dos vectors no són gaire diferents.

```
> rowMeans(P) # mitjana per files

       N       NE        E       SE        S       SW        W       NW
0.294050 0.200625 0.086250 0.122375 0.129175 0.084875 0.052775 0.029875
```

3. Guia de resolució en R. Àlgebra per a ciència de dades

1) Definició i operacions d'una matriu A de dimensions $n \times m$:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix}$$

```
> # A <-matrix(c(...), n,m,byrow= T o F)
> # help("matrix"); example("matrix")
> A <-matrix(c(3,2,1,3), 2,2,byrow= T)
> A

     [,1] [,2]
[1,]    3    2
[2,]    1    3

> det(A)   # determinant de la matriu

[1] 7

> t(A)     # matriu transposada

     [,1] [,2]
[1,]    3    1
[2,]    2    3

> solve(A) # matriu inversa A^(-1)

           [,1]       [,2]
[1,]  0.4285714 -0.2857143
[2,] -0.1428571  0.4285714

> library(expm) # llibreria
> A %^% 2  # matriu al quadrat

     [,1] [,2]
[1,]   11   12
[2,]    6   11

> rankMatrix(A)[1]  # rang de la matriu

[1] 2
```

Nota

El símbol # serveix per escriure comentaris. Els símbols de cert o fals són `T` o `F` i també `TRUE` o `FALSE`.

2) Valors i vectors propis d'una matriu quadrada A: $A\vec{v} = \lambda\vec{v}$, en què λ és el valor propi (VAP) i $\vec{v} = (v_1, \ldots, v_n)$ és el vector propi (VEP):

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

```
> # V <- eigen(A)
> # help("eigen"); example("eigen")
> V <- eigen(A)
> V # valors i vectors propis de la matriu

eigen() decomposition
$values
[1] 4.414214 1.585786

$vectors
          [,1]       [,2]
[1,] 0.8164966 -0.8164966
[2,] 0.5773503  0.5773503

> vaps <- V$values
> abs(vaps) # valors propis en valor absolut

[1] 4.414214 1.585786

> veps <- V$vectors
> # A= veps * diag(vaps) * veps^(-1) descomposició
> diag(vaps); solve(veps)

         [,1]     [,2]
[1,] 4.414214 0.000000
[2,] 0.000000 1.585786

           [,1]      [,2]
[1,]  0.6123724 0.8660254
[2,] -0.6123724 0.8660254

> veps %*% diag(vaps) %*% solve(veps)

     [,1] [,2]
[1,]    3    2
[2,]    1    3

> sum(diag(A)) == sum(vaps) # traça de la matriu = suma dels vaps

[1] TRUE
```

3) Ajust lineal. Recta de regressió $y = ax + b$, en què a i b són coeficients reals.
Mètode dels mínims quadrats (*lm* = *linear model*):

$$\text{Dades} \quad \begin{array}{c|l} X & x_1, x_2, \ \ldots, \ x_n. \\ \hline Y & y_1, y_2, \ \ldots, \ y_n. \end{array}$$

```
> # ajust <- lm(y ~ x) # y=a*x+b
> # summary(ajust) # resum
> # help(" lm"); example(" lm")
> x <- c(3,2,1); y <- c(1,3,-5);
> ajust <- lm(y ~ x)
> coef <- coefficients(ajust); # y=a*x+b, a=coef[2], b=coef[1]
> coef
```

```
(Intercept)           x
  -6.333333    3.000000
```

```
> yy <- fitted(ajust) # valors de les prediccions
> yy
```

```
         1          2          3
 2.6666667 -0.3333333 -3.3333333
```

```
> plot(x,y,col= "blue")    # dades (cercles)
> points(x,yy, pch=2)      # prediccions (triangles)
> abline(ajust,col = "red")  # recta
> title("Recta de regressió")
```

4) Resolució de sistemes d'equacions lineals $A\vec{x} = \vec{b}$:

$$\begin{pmatrix} a_{11} & \ldots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

```
> # help("solve"); example("solve")
> b <- c(3,5)
> cbind(A,b)       # matriu ampliada del sistema
```

```
         b
[1,] 3 2 3
[2,] 1 3 5
```

```
> x <- solve(A,b) # solució del sistema lineal
> x
```

```
[1] -0.1428571  1.7142857
```

5) Cadenes de Markov $\vec{x}_{t+1} = P\vec{x}_t$, $t \geq 0$, en què P és la matriu de les probabilitats de transició (per columnes) i $\vec{x}_t$ és el vector (columna) dels n estats del sistema en el temps t:

$$\begin{pmatrix} x_{t+1}^{(1)} \\ \vdots \\ x_{t+1}^{(n)} \end{pmatrix} = \begin{pmatrix} P_{11} & \dots & P_{1n} \\ \vdots & & \vdots \\ P_{n1} & \dots & P_{nn} \end{pmatrix} \begin{pmatrix} x_t^{(1)} \\ \vdots \\ x_t^{(n)} \end{pmatrix}$$

```
> # Etiquetes <- c("Estat1", "Estat2", ...) # n estats
> # P <- matrix(c(...), n,n, byrow=T o F, dimnames= list(Etiquetes, Etiquetes))
> # matriu transició
> # CM <- new("markovchain", states= Etiquetes, byrow= T o F, transitionMatrix= P,
> #       name= "Títol del problema")
> # plotmat(P, pos= c(n1,n2,n3 ...),
> #         main = "Diagrama d'estats i probabilitats de transició: problema")
> # CM^t                          # cadena de Markov al cap de t unitats de temps
> # Estat estacionari = distribució límit d'estats (%), (columnes o files)
> # 100*steadyStates(t(CM) o CM)
> library(markovchain); library(diagram) # llibreries
> Etiquetes <- c("NYC", "BCN") # 2 estats
> # matriu de transició (per columnes)
>
> P <- matrix(c(0.7,0.3,0.4,0.6),2,2,dimnames= list(Etiquetes, Etiquetes))
> CM <- new("markovchain",states= Etiquetes,byrow= F,
>                      transitionMatrix= P,name= "Moviment treballadors")
> CM # Cadena de Markov

Moviment treballadors
 A  2 - dimensional discrete Markov Chain defined by the following states:
 NYC, BCN
 The transition matrix (by cols) is defined as follows:
    NYC BCN
NYC 0.7 0.4
BCN 0.3 0.6

> plotmat(P, pos= c(2), main =
>               "Diagrama d'estats i probabilitats de transició: moviment treballadors")
> CM^20 # cada columna és la distribució límit d'estats

Moviment treballadors^20
 A  2 - dimensional discrete Markov Chain defined by the following states:
 NYC, BCN
 The transition matrix  (by cols)  is defined as follows:
          NYC       BCN
NYC 0.5714286 0.5714286
BCN 0.4285714 0.4285714
```

```
> 100*steadyStates(t(CM)) # estat estacionari (%) = distribució límit d'estats (%)

          NYC      BCN
[1,] 57.14286 42.85714

> V <- eigen(P); vaps <- V$values; vaps[2]/vaps[1]

[1] 0.3
```